A Distributed Rate Allocation Algorithm for Slepian-Wolf Coding based Data Aggregation in Wireless Sensor Networks

Jun Zheng, Zhenzhong Huang, Qihang Shu, Nathalie Mitton

A Distributed Rate Allocation Algorithm for Slepian-Wolf Coding based Data Aggregation in Wireless Sensor Networks

Jun Zheng[1], Zhenzhong Huang[1], Qihuang Shu[1], and Nathalie Mitton[2]

[1]National Mobile Communications Research Laboratory
Southeast University
Nanjing, Jiangsu 210096, China
[2]Inria Lille – Nord Europe
40 Avenue Halley, 59650 Villeneuve D'ASCQ, France
Email: junzheng@seu.edu.cn

***Abstract*—This paper considers the Slepian-Wolf coding based energy-minimization rate allocation problem in a wireless sensor network (WSN) and propose a distributed rate allocation algorithm to solve the problem. The proposed distributed algorithm is based on an existing centralized rate allocation algorithm which has a high computational complexity. To reduce the computational complexity of the centralized algorithm and make the rate allocation performable in a distributed manner, we make necessary modifications to the centralized algorithm by reducing the number of sets in calculating the average energy consumption cost and limiting the number of conditional nodes that a set can use. Simulation results show that the proposed distributed algorithm can significantly reduce the computational time when compared with the existing centralized algorithm at the cost of the overall energy consumption for data transmission and the total amount of data transmitted in the network.**

***Index Terms*—distributed rate allocation; Slepian-Wolf coding; data aggregation; wireless sensor network**

I. INTRODUCTION

In a wireless sensor network (WSN), a number of sensor nodes are densely deployed in a region of interest to collect data and send the data to the data sink[1]. The data collected by different sensors usually have a certain spatial correlation and the degree of the spatial correlation increases as the distance between sensor nodes decreases[2]. The spatial correlation can lead to considerable data redundancy in the network and transmitting redundant data would cause unnecessary energy consumption. To increase energy efficiency and prolong network lifetime, it is necessary to perform in-network data aggregation to remove the data redundancy in the network.

Slepian-Wolf coding[3] is a distributed source coding technique that can remove data redundancy in a WSN without communication between sensor nodes. When using Slepian-Wolf coding for data aggregation in a WSN, rate allocation is an important issue that has a big impact on the performance of data compression and thus the energy consumption for data transmission in the network [4-5]. To reduce the energy consumption for data transmission, it is desirable to find an optimal rate allocation that can maximize data compression and thus minimize the amount of data transmitted in the network, which is referred to as the Slepian-Wolf coding based energy-minimization rate allocation problem. To solve this problem, Liu et al. has proposed a centralized rate allocation algorithm [6], which, however, requires a high computational complexity and is thus not practical for being applied to a large network. To address this limitation, it is interesting to explore a distributed rate allocation algorithm to solve the problem, which is the motivation for this work.

In this paper, we consider the Slepian-Wolf coding based energy-minimization rate allocation problem in a WSN, and propose a distributed rate allocation algorithm to solve the problem. The proposed distributed algorithm is based on the centralized rate allocation algorithm proposed in [6]. To reduce the computational complexity of the centralized algorithm and make the rate allocation performable in a distributed manner, we make necessary modifications to the centralized algorithm by reducing the number of sets in calculating the average energy consumption cost and limiting the number of conditional nodes that a set can use. Simulation results show that the proposed distributed algorithm can significantly reduce the computational time as compared with the existing centralized algorithm at the cost of the overall energy consumption for data transmission and the total amount of data transmitted in the network.

The rest of the paper is organized as follow. Section II introduces the concept of Slepian-Wolf coding and describes the Slepian-Wolf coding based energy-minimization rate allocation problem considered in this paper. Section III presents the proposed distributed rate allocation algorithm for solving the rate allocation problem. Section IV evaluates the performance of the proposed distributed algorithm through simulation results. Section V concludes this paper.

This work was supported by the National Natural Science Foundation of China under Grant No. 61071115, the Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20110092110007, and the Research Fund of National Mobile Communications Research Laboratory, Southeast University, China, under Grant No. 2012A02.

II. SLEPIAN-WOLF CODING AND PROBLEM STATEMENT

This section introduces the concept of Slepian-Wolf coding and describes the problem considered in this paper.

2.1 Slepian-Wolf Coding

Consider a WSN network with one sink and N sensor nodes that are uniformly distributed in a region of interest, where each node k produces reading X_k. All the readings produced by the sensor nodes constitute a set of jointly stationary ergodic sources, which is denoted by $\boldsymbol{X} = (X_1, X_2, \ldots, X_{\mathrm{N}})$. Assume that the distribution of $\boldsymbol{X} = (X_1, X_2, \ldots, X_{\mathrm{N}})$ is $p(x_1, x_2, \ldots, x_{\mathrm{N}})$, which corresponds to the spatial correlation structure known by each node *a priori*. According to the Slepian-Wolf Theorem [3], the sensor nodes can jointly encode their data without communicating with each other, with a coding rate (in *bits*) lower-bounded by their joint entropy $H(X_1, X_2, \ldots, X_{\mathrm{N}})$ as long as their respective coding rates are under the constraints given by

$$R(\Phi) \geq H(\boldsymbol{X}(\Phi) / \boldsymbol{X}(\Phi^c))$$

for all $\Phi \subseteq \Omega = \{1, 2, \cdots, \mathrm{N}\}$, where $\Omega = \{1, 2, \cdots, \mathrm{N}\}$ is a set of the indices of sensor nodes in the network, Φ^{c} denotes the complementary set of Φ, $H(\boldsymbol{X})$ denotes the entropy of $\boldsymbol{X}$, and

$$R(\Phi) = \sum_{k \in \Phi} R_k \,,$$

$$\boldsymbol{X}(\Phi) = \{X_k, \, k \in \Phi\} \,,$$

$$\boldsymbol{X}(\Phi^{\mathrm{c}}) = \{X_k, \, k \notin \Phi\} \,.$$

For example, for a simple network with two sensor nodes producing readings X_1 and X_2, their individual coding rates should be subject to

$$R_1 \geq H(X_1 / X_2) \,,$$

$$R_2 \geq H(X_2 / X_1) \,,$$

$$R_1 + R_2 \geq H(X_1, X_2) \,.$$

According to Chain Theory [3], under the above constraints, there always exists a rate allocation for the two nodes, which makes the total coding rate (bits) of the two nodes equal to their joint entropy, i.e.,

$$R_1 + R_2 = H(X_1) + H(X_2 / X_1) = H(X_1, X_2) \,.$$

Accordingly, for any order of N nodes, there always exists a rate allocation (vector) $\{R_k\}_{k=1}^{\mathrm{N}}$ that makes the number of bits generated by all the nodes equal to their joint entropy, i.e.,

$$\sum_{k=1}^{\mathrm{N}} R_k = H(X_1, X_2, \cdots, X_{\mathrm{N}}) \,,$$

where $R_1 = H(X_1)$ and $R_k = H(X_k / X_{k-1}, X_{k-2}, \cdots, X_1)$.

2.2 Problem Statement

Consider the same network described in Section 2.1. For the data produced by each sensor node, i.e., $\boldsymbol{X} = (X_1, X_2, \ldots, X_{\mathrm{N}})$, Slepian-Wolf coding is used to jointly encode the data. The encoded data at each sensor node are then transmitted to the sink node along the shortest path between the sensor node and the sink, and are jointly decoded at the sink.

Assume that the coding rate allocated for each sensor node is R_1, R_2, …, R_{N}, respectively. According to the Slepian-Wolf theorem, the data can be jointly decoded without loss at the sink subject to the following constraint:

$$R(\Phi) \geq H(\boldsymbol{X}(\Phi) / \boldsymbol{X}(\Phi^c)) \tag{1}$$

for $\Phi \subseteq \Omega = \{1, 2, \cdots, \mathrm{N}\}$, where

$$R(\Phi) = \sum_{k \in \Phi} R_k \,,$$

$$\boldsymbol{X}(\Phi) = \{X_k, \, k \in \Phi\} \,,$$

$$\boldsymbol{X}(\Phi^{\mathrm{c}}) = \{X_k, \, k \notin \Phi\} \,.$$

Under the above assumptions, the problem we consider is to find an optimal rate allocation for each sensor node to minimize the amount of data transmitted in the network and thus the overall energy consumption for data transmission subject to the constraint given in Eq. (1).

To further describe the above problem, we assume that the path cost for node k to send its encoded data to the sink along the shortest path is w_k. The coding rate allocated to node k is R_k. Thus, the energy consumption for node k to transmitting the encoded data with rate R_k is

$$E(R_k) = w_k \times c(R_k) \,, \tag{2}$$

where $c(\cdot)$ is a function only related to the amount of data or the coding rate. In this paper, we assume that

$$c(R_k) = \exp(R_k) \,. \tag{3}$$

Under the above assumptions, the overall energy consumption for transmitting the encoded data to the sink in the network is

$$E = \sum_{k=1}^{\mathrm{N}} E(R_k) = \sum_{k=1}^{\mathrm{N}} w_k \times \exp(R_k) \,. \tag{4}$$

Therefore, the problem is to find a rate allocation $\boldsymbol{R}^* = (R_1^*, R_2^*, \ldots, R_{\mathrm{N}}^*)$ for each sensor node so that the overall energy consumption for data transmission in the network is minimized, i.e.,

$$\boldsymbol{R}^* = (R_1^*, \cdots, R_{\mathrm{N}}^*) = \arg[\min \sum_{k=1}^{\mathrm{N}} w_k \times \exp(R_k)] \,, \tag{5}$$

subject to

$$R(\Phi) \geq H(\boldsymbol{X}(\Phi) \,|\, \boldsymbol{X}(\Phi^c)) \,.$$

2.3 Centralized Rate Allocation Algorithm

To solve the above problem, Liu et al. have proposed a centralized rate allocation algorithm in [6], which is based on the water-filling algorithm and gives an optimal solution to the problem.

In this algorithm, all sensor nodes in the network are divided into two node sets: conditional node set Φ_1, which is null initially, and non-conditional node set Φ_2, where $\Phi_2 = \Omega / \Phi_1$. In each round, it selects a non-null subset that meets a given condition from all the subsets of Φ_2, adds all the nodes in the selected subset into the conditional set Φ_1, and then removes these nodes from the non-conditional set Φ_2. This procedure is repeated until the non-conditional node set Φ_2 becomes null.

The main procedure of the centralized rate allocation algorithm is described as follows:

1) Let $n = \mathrm{N}$, $\Phi_1 = \phi$, where n is the number of nodes in Ω / Φ_1;
2) Calculate the value of $H(\boldsymbol{X}(\Phi) \mid \boldsymbol{X}(\Phi_1))$, where Φ is any subset of set Ω / Φ_1 and the number of possible Φs is $(2^n - 1)$;
3) Select a set Φ_0 from all possible Φs, which has the minimum node's average energy consumption cost value, i.e.,

$$\Phi_0 = \arg[\min_{\Phi \in \boldsymbol{\Omega} / \Phi_1} \{H(\boldsymbol{X}(\Phi) \mid \boldsymbol{X}(\Phi_1)) + \sum_{j \in \Phi} \ln w_j\} / |\Phi|] ; \quad (6)$$

4) For each node k in set Φ_0, set the rate R_k, i.e.,

$$R_k = \{H(\boldsymbol{X}(\Phi_0) \mid \boldsymbol{X}(\Phi_1)) + \sum_{j \in \Phi_0} \ln w_j\} / |\Phi_0| - \ln w_k ; \quad (7)$$

5) Update $n = n - |\Phi_0|$, $\Phi_1 = \Phi_1 \cup \Phi_0$; If $n = 0$, stop; otherwise, go to step 2).

The above algorithm is a global rate allocation algorithm which gives an optimal solution to the problem in Eq. (5). However, it has two weaknesses:

(a) Step 2) introduces a high computational complexity;

(b) Step 3) requires a global comparison of all the values obtained in step 2).

Due to the above two weaknesses, the algorithm has a high computational complexity and thus can only be performed in a centralized manner.

III. Distributed Rate Allocation Algorithm

In this section, we present our proposed distributed rate allocation algorithm for Slepian-Wolf coding based data aggregation in WSNs. We first make necessary modifications to address the two weaknesses of the centralized algorithm and then describe the procedure of the proposed distributed algorithm.

3.1 Reduction of Computational Complexity

For the centralized rate allocation algorithm, it needs to calculate the average energy consumption cost value for any possible subset Φ of set Ω / Φ_1, i.e.,

$$\{H(\boldsymbol{X}(\Phi) \mid \boldsymbol{X}(\Phi_1)) + \sum_{j \in \Phi} \ln w_j\} / |\Phi| . \quad (8)$$

Since the number of nodes in Ω / Φ_1 is *n*, the number of possible Φs is $(2^n - 1)$. This means that the computational time of the algorithm increases exponentially with the number of nodes in the network. Obviously, when *n* becomes large, it is difficult for a sensor node to undertake the huge amount of computation because of the limitation in its computational capability.

To reduce the computational complexity, it is necessary to reduce the number of Φs that need to be calculated in step 2). By observing step 3), it is found that the purpose of step 2) is to find the set Φ_0 with the minimum average energy consumption cost value. Thus, we only need to calculate those Φs that lead to a smaller value of the average energy consumption cost, and ignore other Φs. In this way, the amount of computational time can be largely reduced. However, this should not significantly degrade the performance of the algorithm. That is, the algorithm should still be able to find the set Φ_0 with the minimum average energy consumption cost value. Therefore, the next question is "What Φs should be selected for calculation in step 2)?"

As mentioned earlier, Eq. (8) represents the average energy consumption cost of set Φ, which consists of two parts: the first part is the average conditional entropy of set Φ and the second part is proportional to the average path cost of set Φ. In general, the data correlation degree between different nodes decreases with the increase of the distance between the nodes. For this reason, the average entropy value of a set whose nodes are away from each other is usually larger than that of a set whose nodes are closer to each other. Based on this fact, it is unnecessary to calculate the cost values of those sets whose nodes are away from each other in step 2). In another word, we only need to calculate those subsets whose nodes are closer to each other. The specific definition of those subsets that need to be calculated will be given in the next subsection.

3.2 Distributed Implementation of Rate Allocation

As explained in Section 2.3, the centralized algorithm has two weaknesses that make it only feasible to be performed in a centralized manner:

(a) Step 2) needs to calculate the conditional entropy $H(\boldsymbol{X}(\Phi) \mid \boldsymbol{X}(\Phi_1))$ of any possible subset Φ of set Ω / Φ_1 in step 2). For each subset Φ, it must know the information on all conditional nodes, which is usually globally distributed and thus makes the calculation infeasible to be performed in a distributed manner.

(b) Step 3) needs to compare all the conditional entropy values obtained in step 2), which is a global comparison and thus can be only performed in a centralized manner.

To make the algorithm implementable in a distributed manner, the above two weaknesses must be addressed.

A) Limitation of Conditional Nodes

As discussed earlier, to make the computational process of the centralized algorithm implementable in a distributed manner, it is necessary to impose some limitation on the conditional nodes that a set Φ can use. Specifically, we can limit the conditional nodes a set can use to only its local nodes. That is, the conditional nodes of a particular node must be one-hop neighbor nodes of that node. In addition to this limitation, a conditional node must also meet another condition, i.e., it must be a conditional node selected in the last round of calculation (or in set Φ_1) like the centralized algorithm.

Next we discuss how to determine the conditional nodes that a set Φ can use. The notations used in the discussion are defined as follows.

$P_k^{'}$: a set consisting of all one-hop neighbor nodes of node *k*;

Φ_1: The conditional node set of the network. The conditional node set of a particular set Φ is a subset of Φ_1;

$I(\cdot)$: the entropy of the mutual information between different nodes.

Φ_{1k}: the conditional node set of node *k*.

$\Phi_{1k}^{'}$: the conditional node set of set $P_k^{'}$, where $\Phi_{1k}^{'} = \{\Phi_{1k_i}, k_i \in P_k^{'}\}$;

P_k: a node set that belongs to set $P_k^{'}$ but does not belong to set $\Phi_{1k}^{'}$, i.e., $P_k = \mathrm{P}_k^{'} / \Phi_{1k}^{'}$.

We know that the entropy value of mutual information between two nodes has a negative correlation to the distance between the two nodes, and we have

$$H(\boldsymbol{X}(\Phi) \mid \boldsymbol{X}(\Phi_1)) = H(\boldsymbol{X}(\Phi)) - I(\boldsymbol{X}(\Phi), \boldsymbol{X}(\Phi_1)) \tag{9}$$

$$I(\boldsymbol{X}(\Phi), \boldsymbol{X}(\Phi_1)) = I(\boldsymbol{X}(\Phi_1^{'}), \boldsymbol{X}(\Phi)/\boldsymbol{X}(\Phi_1^{''})) + I(\boldsymbol{X}(\Phi_1^{''}), \boldsymbol{X}(\Phi)/\boldsymbol{X}(\Phi_1^{'})) \tag{10}$$

$$\Phi_1 = \Phi_1^{'} + \Phi_1^{''}$$

where $\Phi_1^{'}$ denotes the set consisting of the conditional nodes which are closer to set Φ, and $\Phi_1^{''}$ denotes the set consisting of the conditional nodes which are away from set Φ, $I(\cdot)$ denotes the entropy of the mutual information between different nodes.

According to the relevant properties of mutual information, the value of $I(\boldsymbol{X}(\Phi_1^{''}), \boldsymbol{X}(\Phi)/\boldsymbol{X}(\Phi_1^{'}))$ is relatively much smaller than that of $I(\boldsymbol{X}(\Phi_1^{'}), \boldsymbol{X}(\Phi)/\boldsymbol{X}(\Phi_1^{''}))$, and thus can be neglected. Similarly, we have

$$\begin{aligned}
&I(\boldsymbol{X}(\Phi_1^{'}), \boldsymbol{X}(\Phi)/\boldsymbol{X}(\Phi_1^{''})) \\
&= H(\boldsymbol{X}(\Phi_1^{'})/\boldsymbol{X}(\Phi_1^{''})) - H(\boldsymbol{X}(\Phi_1^{'})/\boldsymbol{X}(\Phi_1^{''}), \boldsymbol{X}(\Phi)) \\
&= H(\boldsymbol{X}(\Phi_1^{'})) - H(\boldsymbol{X}(\Phi_1^{'})/\boldsymbol{X}(\Phi)) \\
&= I(\boldsymbol{X}(\Phi_1^{'}), \boldsymbol{X}(\Phi))
\end{aligned} \tag{11}$$

Thus, we can replace $I(\boldsymbol{X}(\Phi_1^{'}), \boldsymbol{X}(\Phi)/\boldsymbol{X}(\Phi_1^{''}))$ with $I(\boldsymbol{X}(\Phi_1^{'}), \boldsymbol{X}(\Phi))$. According to Eq. (9) - Eq. (11), we have

$$\begin{aligned}
H(\boldsymbol{X}(\Phi)/\boldsymbol{X}(\Phi_1)) &= H(\boldsymbol{X}(\Phi)) - I(\boldsymbol{X}(\Phi), \boldsymbol{X}(\Phi_1)) \\
&= H(\boldsymbol{X}(\Phi)) - I(\boldsymbol{X}(\Phi), \boldsymbol{X}(\Phi_1^{'})) \\
&= H(\boldsymbol{X}(\Phi)/\boldsymbol{X}(\Phi_1^{'})).
\end{aligned}$$

Therefore, we can calculate the value of $H(\boldsymbol{X}(\Phi)/\boldsymbol{X}(\Phi_1^{'}))$ instead of the value of $H(\boldsymbol{X}(\Phi)/\boldsymbol{X}(\Phi_1))$. That is, for each set Φ, we select a set of conditional nodes $\Phi_1^{'}$ from condition nodes Φ_1. This means that for each set Φ its conditional node set is no longer Φ_1, but set $\Phi_1^{'}$, which belongs to Φ_1 and consists of only those nodes closer to set Φ. For simplicity, we only consider those "closer" nodes as conditional nodes. Specifically, for a particular set Φ, a node in its conditional node set $\Phi_1^{'}$ must meet the following condition: a one-hop neighbor node of a node in set Φ. Of course, we can consider a larger range (e.g., two hops) for selecting the conditional nodes. Obviously, the larger the range, the more conditional nodes for a particular node, and the less the performance degrades.

Based on the above definition on conditional nodes, we now discuss how to determine the conditional node set $\Phi_1^{'}$ for set Φ.

For a particular node *k*, its conditional node must meet the conditions below:

1) The node belongs to Φ_1;
2) The node is one-hop away from node *k*.

In this case, for set $P_k^{'}$, its conditional node set is $\Phi_{1k}^{'} = \{\Phi_{1k_i}, k_i \in P_k^{'}\}$. According to the definition of set P_k, we can obtain $P_k = \mathrm{P}_k^{'} / \Phi_{1k}^{'}$.

Based on the above analysis, for a particular node *k*, we only need to calculate the average energy consumption cost value of a set that consists of node *k* and any number of its one-hop neighbor nodes, and the nodes in that set cannot be a conditional node in set Φ_1. Therefore, P_k and its real subsets constitute all sets that need to be calculated for node *k*. In this case, the number of sets that need to be calculated becomes $\sum_{j=1}^{\mathrm{n}} 2^{|\boldsymbol{P}_k|}$. Obviously, we have $\sum_{j=1}^{\mathrm{n}} 2^{|\boldsymbol{P}_k|} << 2^{\mathrm{n}} - 1$ when *n* is large. As a result, the computational complexity of the distributed algorithm is largely reduced.

B) Reduction of Comparison Scope

Next let us consider the global comparison problem in step 3). To make the comparison performable in a distributed manner, it is necessary to reduce the comparison scope in step 3).

As explained earlier, the purpose of the global comparison in step 3) is to find a set with the minimum average energy consumption cost among all possible sets. For this purpose, we can further divide the global comparison into two steps.

(1) For each node *k*, compare the average energy consumption cost values of all sets that have node *k* as the center and find the set with the minimum value, which is called representative set of node *k*.

(2) For each node, compare the average energy cost value of its representative set and find the one with the minimum value.

Since step (2) is a global comparison, it is necessary to reduce the comparison scope in this step. By careful observation, we can find the following two facts:

(1) For a particular node *v*, only when the set consisting of the nodes within the two hops of node *v* is selected in the current round will the average energy consumption cost value of the representative set of node *v* change or need to be recalculated in the next round of calculation.

(2) For a particular node *v*, if the average energy consumption cost of its representative set is smaller than that of the representative set of any of its neighbor nodes within two hops, its representative set is selected and added to the conditional node set earlier than the representative set of any of its neighbor nodes within two hops.

Based on the above facts, we can obtain the following conclusion: for a particular node *v*, we only need to consider the neighbor nodes within its two hops, and can ignore those nodes that are not within its two hops. Therefore, in the distributed algorithm, for those sets that have a particular node *v* as their center node (i.e., all the other nodes in a set are one-hop neighbor nodes of node *v*), a set that needs to be compared with these sets must meet the following conditions:

1) The set has a particular node *u* as its center node and the other nodes in the set are one-hop away from node *u*;
2) The center node, node *u*, is at most two hops away from node *v*.

As a result, the global comparison in the centralized algorithm can be replaced by the local comparison within a two-hop range of a particular node, which makes it feasible to perform rate allocation in a distributed manner. It should be noted that since the algorithm is performed in a distributed manner, in each round of calculation, more than one set may be selected, which is a difference compared with the centralized algorithm.

3.3 Distributed Rate Allocation Algorithm

After making the modifications to the centralized algorithm, we now present the proposed distributed rate allocation algorithm and describe its major procedure as follows:

1) Let $n = \mathrm{N}$, $\Phi_1 = \phi$, where n is the number of nodes in Ω / Φ_1;

2) For each node k in Ω / Φ_1, select the representative set of node k, which satisfies the following condition:

$$\boldsymbol{P}_{ck} = \arg[\min_{\boldsymbol{P}_{sk} \subseteq \boldsymbol{P}_k} \{\mathrm{H}(\boldsymbol{X}(\boldsymbol{P}_{sk}) / \boldsymbol{X}(\boldsymbol{C}(\boldsymbol{P}_{sk}))) + \sum_{j \in \boldsymbol{P}_{sk}} \ln w_j\} / |\boldsymbol{P}_{sk}|]$$

where $\boldsymbol{C}(\boldsymbol{P}_{sk})$ denotes the conditional node set of set $\boldsymbol{P}_{sk}$;

3) For each node k in Ω / Φ_1, comparing the average cost value of its representative set $\boldsymbol{P}_{ck}$ with that of the representative set $\boldsymbol{P}_{cl}$ of its neighbor node l within two hops. If $\boldsymbol{P}_{ck}$ has the minimum average cost value, for each node $k_i \in \boldsymbol{P}_{ck}$, its rate is set to

$$\boldsymbol{R}_{k_i} = \{\mathrm{H}(\boldsymbol{X}(\boldsymbol{P}_{ck}) / \boldsymbol{X}(\boldsymbol{C}(\boldsymbol{P}_{ck}))) + \sum_{j \in \boldsymbol{P}_{sk}} \ln w_j\} / |\boldsymbol{P}_{ck}| \text{-} \ln w_k ;$$

Then, let $\Phi_1 = \Phi_1 \bigcup \boldsymbol{P}_{ck}$, $n = n - |\boldsymbol{P}_{ck}|$;

4) If n=0, stop. Otherwise, go to step 2).

In the above algorithm, each step only needs local information, which makes it feasible for the algorithm to be performed in a distributed manner and significantly reduces the computational complexity of the algorithm.

IV. SLIMULATION RESULTS

In this section, we evaluate the performance of the proposed distributed rate allocation algorithm through simulation results. For evaluation, we compare the distributed algorithm with the centralized rate allocation algorithm in terms of the overall energy cost for data transmission, total amount of data transmitted in the network, and total amount of computational time.

Given that the centralized algorithm has a high computational complexity, it is inappropriate to consider a network with a large number of nodes in the simulation experiments. For this reason, we considered a 100m×100m area with 100 nodes uniformly distributed. We used the energy model in [7]. and the correlation model in [8]. We assumed that the amount of data generated by one node without compression is 1 bit. The path used for data transmission is the shortest path tree, and the path cost w_k of node k is d_k^2, where d_k is the path distance between node k and the sink.

Fig. 1 and Fig. 2 show the overall energy consumption and the total amount of data transmitted in the network with the centralized algorithm and the distributed algorithm, respectively. It is seen that the performance of the distributed algorithm degrades as compared with that of the centralized algorithm. This is because in the distributed algorithm the conditional information a particular node can use is limited to only local information. In this case, the conditional entropy value of the node will increase. Since the coding rate of a node is equal to the conditional entropy of the node, the coding rate (or the amount of data) would also increase, which results in the increase of the overall energy consumption for data transmission in the network. On the other hand, the value of the performance degradation depends on the data correlation degree between different nodes. The larger the data correlation degree, the larger the performance degrades. This is because in the distributed algorithm we only consider the nodes within one hop of a node, which would result in less conditional information that can be used by a node.

Fig. 3 shows the computational time in terms of simulation time with the centralized algorithm and the distributed algorithm, respectively. It is seen that the distributed algorithm significantly reduces the computational time as compared with the centralized algorithm.

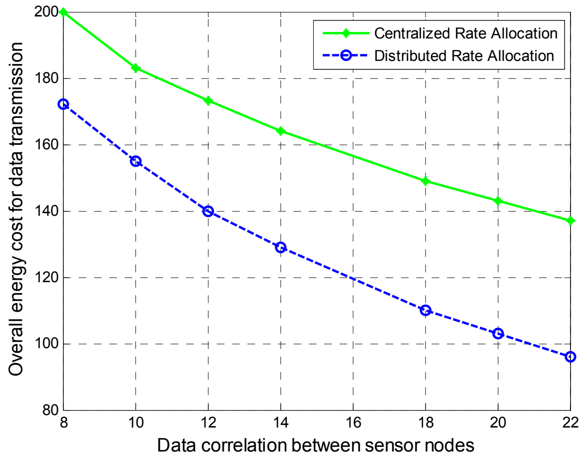

Fig.1 Overall energy cost in the network.

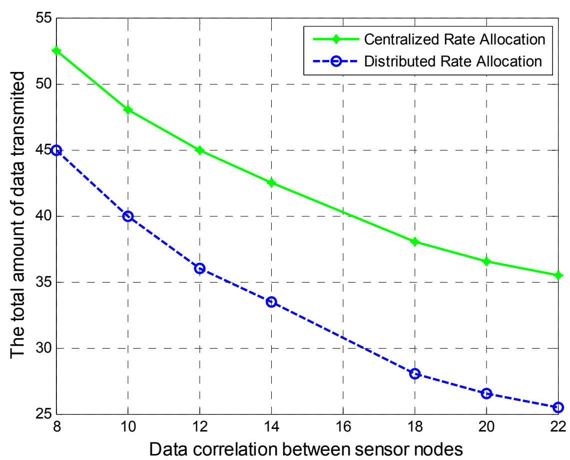

Fig.2 Total amount of data transmitted in the network.

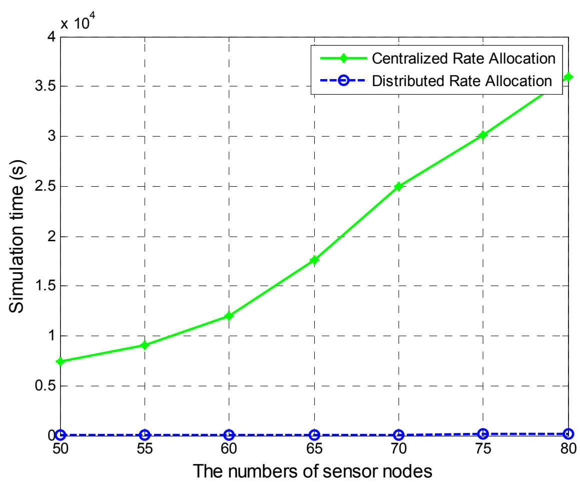

Fig.3 Simulation time.

V. CONCLUTIONS

In this paper, we considered the Slepian-Wolf coding based energy-minimization rate allocation problem in a WSN and proposed a distributed rate allocation algorithm to solve the problem. The proposed distributed algorithm is based on the existing centralized algorithm proposed in [6], which has a high computational complexity. To reduce the computational complexity and make the rate allocation performable in a distributed manner, we make necessary modifications to the centralized algorithm by reducing the number of sets in calculating the average energy consumption cost and limiting the number of conditional nodes that a set can use. The simulation results show that the proposed distributed algorithm can significantly reduce the computational time as compared with the existing centralized algorithm at the cost of the overall energy consumption for data transmission and the total amount of data transmitted in the network.

REFERENCES

[1] J. Zheng and A. Jamalipour, Wireless Sensor Networks: A Networking Perspective, Wiley IEEE Press

[2] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York, NY, USA: John Wiley and Sons, Inc., 1991.

[3] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” IEEE Transactions on Information Theory, vol. 19, no. 4, July 2003, pp. 471–480.

[4] M. C Vuran, O. B. Akan, and I. F. Akyildiz, “Spatio-temporal correlation: theory and application for wireless sensor network,” Elsevier Computer Networks, vol. 45, no. 3, Mar. 2004, pp. 245–259.

[5] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, “Networked Slepian-Wolf: theory, algorithms, and scaling laws,” IEEE Transactions on Information Theory, vol. 51, no. 12, Nov. 2005, pp. 4057– 4073.

[6] W. Liu, L. Dong, and W. Zeng, “Power efficient rate allocation for Slepian-Wolf coding over wireless sensor networks ,” in Proc of 2008 ICASSP, Mar. 2008, pp. 2713–2716.

[7] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, “An application-specific protocol architecture for wireless microsensor network,” IEEE Transactions on Wireless Communications, vol. 1, no. 4, Oct. 2002, pp. 660–671.

[8] S. Pattem, B. Krishnamachari, and R. Govindan, “The impact of spatial correlation on routing with compression in wireless sensor networks,” in Proc. of 2004 Third International Symposium on Information Processing in Sensor Networks (IPSN’04), Berkeley, CA, Apr. 2004, pp. 28–35.